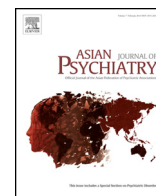


Contents lists available at ScienceDirect

Asian Journal of Psychiatry

journal homepage: www.elsevier.com/locate/ajp

Teacher ratings of ODD symptoms: Measurement equivalence across Malaysian Malay, Chinese and Indian children



Rapson Gomez*

University of Ballarat, Australia

ARTICLE INFO

Article history:

Received 1 July 2013

Accepted 28 October 2013

Keywords:

Oppositional defiant disorder

Cross-cultural equivalence

Malaysia

Multiple-group CFA

ABSTRACT

Background: The study examined the measurement equivalence for teacher ratings across Malaysian Malay, Chinese and Indian children.

Methods: Malaysian teachers completed ratings of the ODD symptoms for 574 Malay, 247 Chinese and 98 Indian children.

Results: The results supported the equivalences for the configural, metric, and error variances models, and the equivalences for ODD latent variances and mean scores.

Discussion: Together, these findings suggest good support for measurement and structural equivalences of the ODD symptoms across these ethnic groups. The theoretical and clinical implications of the findings for cross-cultural equivalence of the ODD symptoms are discussed.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

According to the Diagnostic and Statistical Manual (DSM-IV-TR, [American Psychiatric Association, APA, 2000](#)), oppositional defiant disorder (ODD) is one of the most common childhood psychological disorders. The essential feature of ODD, as described by DSM-IV-TR, is a persistent pattern of negativistic, hostile, and defiant behavior. DSM-IV-TR has a list of eight symptoms for its diagnosis. Despite the fact that the influence of ethno-cultural factors on ODD has not been adequately addressed ([Gomez et al., 2008](#)), an examination of the existing literature shows that the same set of ODD symptoms and criteria are used cross-culturally ([Canino et al., 2010](#)). This practice assumes that there is measurement equivalence (or measurement invariance) for the ODD symptoms across different cultural groups. Measurement equivalence does not refer to observed scores (such as item or mean scores) being the same across different groups. It refers to the same observed scores having the same level of the underlying latent trait. The current study examined the measurement equivalence of teacher ratings of the ODD symptoms across Malaysian Malay, Chinese and Indian children. Thus the question addressed in this study was “Do Malay, Chinese and Indian children with the same latent scores for the ODD symptoms have

the same observed scores?” Demonstration of this would imply measurement equivalence for the ODD symptoms across these groups.

Culture has been defined by [Hofstede \(1980\)](#) in terms of the collective programming of the mind that distinguishes the members of one group of people from another group. Central to this distinction are values, attitudes, beliefs and practices. Such a view recognizes cultural homogeneity of different ethnic groups within the same nation, as well as cultural heterogeneity within the same nation. As pointed out by Hofstede, nations can have many cultures, and cultures can bridge more than one nation. Malaysia is a good example of a nation with multiple cultures, as defined by Hofstede. The Malaysian population comprises 50.5% Malays, 11% Indigenous, 24.2% Chinese, 7.1% Indians, and 1.3% others ([Department of Statistics Malaysia, 2003](#)), and the Malaysian Government recognizes these groups as distinct ethno-cultural groups (emic).

Specific to cultural differences among Malay, Chinese and Indian communities that has implications for child oppositional behaviors, [Quah \(2004\)](#) found that Chinese parents are likely to be more punitive and less warm and involved compared to Malay and Indian parents. Since such parenting practices are associated with oppositional responses among children ([Patterson et al., 1992](#); [Stormshak et al., 2000](#)), it could be speculated that Chinese children would be more oppositional than Malay children and Indian children. If so, this could differentially influence perceptions of oppositional behaviors across these groups of children that could contribute to lack of measurement equivalence for ODD symptoms when they are reported on.

* Correspondence to: School of Health Sciences, University of Ballarat, University Drive, Mt Helen, PO Box 663, Ballarat, VIC 3353, Australia. Tel.: +61 3 5327 6087; fax: +61 3 5327 6087.

E-mail address: rapson.gomez@ballarat.edu.au

Measurement equivalence deals with whether the expected scores on a measure are the same or different across different groups when the groups have the same level of the underlying latent trait scores (Reise et al., 1993). Equivalence is inferred when the expected scores are the same, whereas nonequivalence is inferred if this is not the case. Equivalence would mean that the measure used is measuring conceptually similar constructs across the groups being compared, whereas nonequivalence would mean that the measure used is measuring conceptually different constructs across the groups. If there is weak or no support for measurement equivalence, then it follows that the groups in question cannot be accurately compared in terms of observed scores on the relevant measure. Expressed differently, the same observed score for the groups do not reflect the same level of the underlying trait.

Since the same set of ODD symptoms is used in the same way across different cultures (Canino et al., 2010), it follows that there needs to be cross-cultural measurement equivalence for the ODD symptoms for this practice to have creditability. To date only one study by Gomez et al. (2008) has examined cross-cultural measurement equivalence. For parent ratings of the ODD symptoms, it reported equivalence across ratings of Malaysian and Australian children, and also Malaysian Malay and Chinese children. To date, there has been no study of cross-cultural equivalence of the ODD symptoms for teacher ratings. Since teachers are valuable informants for diagnosis, such data will be useful clinically. Thus the aim of the current study was to examine cross-cultural measurement equivalence for teacher ratings of the ODD symptoms. This was examined across Malaysian Malay, Chinese and Indian children.

2. Methods

2.1. Participants

Teachers rated 574 Malay (255 boys and 319 girls), 247 Chinese (115 boys and 132 girls) and 98 Indian (57 boys and 41 girls) children from the State of Johor in Malaysia. The children were from fourteen randomly selected schools and were between 6 and 12 years of age. The mean ages (*SD*) for Malay, Chinese and Indian children were 8.82 (1.69), 9.32 (1.61) and 8.73 (1.68) years, respectively. The groups differed for age, $F(2, 916) = 8.80, p < .001$, with Chinese children being older than both Malay and Indian children. However the effect sizes for the differences between Chinese and Malay children ($d = 0.30$), and Chinese and Indian children ($d = 0.36$) were both small, based on Cohen's (1992) guidelines (small $\leq .20$, medium $\geq .50$, and large $\geq .80$). There was little difference between Malay and Indian children. In terms of educational attainment, almost all parents had attended or completed secondary education.

2.2. Measure

All teachers completed the Disruptive Behavior Rating Scale (DBRS; Barkley and Murphy, 1998). This rating scale lists the DSM-IV-TR symptoms for ADHD, ODD, and conduct disorder. The focus in the current study is on the ODD symptoms. The 8 ODD symptoms are presented, with the word “often” excluded. Respondents rate the occurrence of each symptom over the past 6-months on a 4-point scale (0 = “never or rarely”, 1 = “sometimes”, 2 = “often”, or 3 = “very often”). The internal reliability of the ODD scale for the study was high at .89. Measurement equivalence was tested using binary scores. For each symptom, the first two response options of the original responses (i.e., options 0 and 1) were recoded as the symptom being absent, while the next two response options (i.e., options 2 and 3) were recoded as the

symptom being present. This type of scoring aligns with the way ODD symptoms are coded for diagnosis.

2.3. Procedure

Ethical approval for the study was obtained from the relevant authorities and bodies (University of Ballarat Human Research Ethics Committee, Federal Ministry of Education of Malaysia, the Ministry of Education of the State of Johor, and the principals of the schools involved). Following this, classroom teachers were issued with the appropriate number of large sealed envelopes to be forwarded to parents, through their students. Each envelope contained a plain language statement or letter providing the background of the study, and a return envelope. Parents were requested to provide the child's age, gender and ethnic background, and their willingness to have the DBRS completed by their children's class teachers. In all, about 1020 parents consented to this. A total of 934 completed the DBRS. Apart for ratings for Malay, Chinese and Indian children, there were ratings for 15 children from other ethnic groups. Excluding this group, the total number of Malay, Chinese and Indian children rated by teachers was 919.

Teachers were provided both English and Malay versions of the DBRS, with instructions to select any one version. The Malay version was developed via forward and backward translation by experts in both languages. A journalist made the forward translation to the Malay language, while a language teacher made the backward translation to English. Following this, the differences between the backward translated version and the original English version as they related to the initial Malay version were resolved by discussion involving both these individuals and a trainee clinical psychologist competent in both Malay and English. This resulted in the final Malay version that was used in the study.

2.4. Statistical procedures

All the CFA analyses in the study were conducted using *Mplus* (Version 6) software (Muthen and Muthen, 2010), and they used the mean and variance-adjusted weighted least squares, or WLSMV. This is a robust estimator for ordered-categorical scores. At the statistical level, model fit was ascertained using the robust WLSMV χ^2 fit function. Non-significant values indicate good fit. However as χ^2 are highly sensitive to sample size and will often show significant values even when there is good fit, the fit is generally evaluated using several approximate fit indices. The approximate fit indices used here were the root mean square error of approximation (RMSEA) and the comparative fit index (CFI). Guidelines suggested are that RMSEA values close to 0.06 or below and CFI values close to .95 or above be taken as indicating good fit (Hu and Bentler, 1998).

Measurement equivalence was tested using the procedure proposed by Millsap and Yun-Tein (2004) for the WLSMV estimator with theta parameterization. Equivalence was tested progressively for configural equivalence (M1), metric equivalence (M2), and error variances equivalence (M3), followed by equivalence for latent factor variances (M4), and mean scores (M5). The test for threshold invariance is not possible with binary scores. The various CFA invariance models were compared using the WLSMV χ^2 difference test that was computed using the option available in *Mplus*.

3. Results

3.1. Functional equivalence of the ratings of ODD symptoms in the Malay and English versions of the DBRS

Initially, the measurement equivalence of for the ODD symptoms rated for the Malay and English versions of the DBRS

Table 1

Results of tests for equivalence across Malaysian Malay, Chinese and Indian children.

| Model (M) | WLSMV χ^2 | df | RMSEA | CFI | ΔM | Δdf | $\Delta WLSMV\chi^2$ |
|--|----------------|----|-------|------|------------|-------------|----------------------|
| M1: configural equivalence | 71.35 | 60 | .025 | .999 | | – | – |
| M2: full metric equivalence (M1 with equality constrained on all factor loadings) | 87.86 | 74 | .025 | .999 | M2–M1 | 14 | 18.08 |
| M3: full error variance equivalence (M2 with equality constrained on error variances) | 92.33 | 88 | .013 | .999 | M3–M2 | 14 | 9.28 |
| M4: equivalence for the variances for ODD latent factors (M2 with equality constrained on ODD factor variances) | 84.38 | 76 | .019 | 1.00 | M4–M2 | 2 | 0.76 |
| M5: equivalence for the latent mean scores for ODD latent factors (M2 with equality constrained on ODD latent mean scores) | 85.69 | 78 | .020 | .999 | M5–M2 | 2 | 0.86 |

Note. WLSMV χ^2 =weighted least square with mean and variance adjusted chi-square; RMSEA=root mean square error of approximation; CFI=comparative fit index. All WLSMV χ^2 and $\Delta WLSMV\chi^2$ values were not significant.

was examined. Of the 934 DBRS ratings completed by teachers, 732 were completed in Malay, and 202 were completed in English. Although details are not provided here (available from the author), the findings showed full measurement equivalence (configural, metric, and error variance). As different language versions can be considered functionally equivalent when they measure the same constructs (Butcher, 1996), these findings can be taken as supportive of the functional equivalence of the different version.

3.2. Evaluation of fit of the one-factor ODD model for ratings of Malay, Chinese and Indian children

The fit of the one-factor ODD model for ratings of Malay, Chinese and Indian children were evaluated separately. The fit indices for ratings of Malay children were χ^2 ($df=20$)=41.27, $p<.01$, RMSEA=.043, and CFI=.998. These were χ^2 ($df=20$)=18.18, ns , RMSEA=.000, and CFI=1.00 for Chinese children. For Indian children they were χ^2 ($df=20$)=19.64, ns , RMSEA=.000, and CFI=1.00. Thus teacher ratings of Malay, Chinese and Indian children showed good model fit.

3.3. Multiple group CFA analyses for measurement equivalence across the cultural groups

Table 1 shows the results of the analyses for equivalence testing across Malaysian Malay, Chinese and Indian children. As shown, the configural model (M1 in the table) showed good fit, thereby supporting the configural equivalence model. The configural model (M1 in the table) and the metric equivalence model (M2 in the table), and also the metric equivalence model (M2 in the table) and the error variances equivalence models (M3 in the table) did not differ from each other. These findings indicate support for the metric equivalence and error variances equivalence models, respectively. The table also shows that there was no difference between the latent factor variances equivalence model (M4 in the table) and the metric equivalence model (M2 in the table). This finding indicates support for equivalence of latent factor variances. There was also no difference between the latent factor mean score equivalence model (M5 in the table) and the metric equivalence model (M2 in the table), thereby supporting the equivalence of latent mean scores. Overall, therefore, there was support for equivalence for all measurement and structural parameters.

4. Discussion

The current study used multiple-group CFA to examine equivalence for teacher ratings of the ODD symptoms across Malaysian Malay, Chinese and Indian children. The findings indicated support for configural equivalence, metric equivalence, error variances equivalence, and equivalence for latent variances and mean scores. Thus, the ratings across these groups were

identical for the factor underlying the observed symptoms (configural equivalence), the strengths of the associations of all the symptoms with its latent factor (metric equivalence), the uniqueness in the symptoms (error variances equivalence), the variance for latent factors (latent variance equivalence), and severity of ODD (same latent mean scores). Overall the findings suggest measurement equivalence for the ODD symptoms and the same mean scores for the ODD factor across teacher ratings of Malaysian Malay, Chinese and Indian children.

The support for equivalence in the ODD symptoms in the current study is consistent with the results of the previous study by Gomez et al. (2008). However, unlike the previous study that examined parent ratings, the current study examined teacher ratings. Also, unlike the previous study that examined the original ordinal scores, the current study examined binary scores. Given that binary scores correspond with the way ODD symptoms are scored for diagnosis, the findings here are more relevant for clinical practice than the previous study by Gomez et al. (2008). Also, unlike the previous study that examined equivalence for rating of Malay and Chinese children, the current study examined equivalence across Malay, Chinese and Indian children.

The findings in this study and the previous study by Gomez et al. (2008) have implications for DSM-IV ODD diagnostic criteria. They raise the possibility that there will be equivalence for the ODD symptoms across different racial, cultural and national groups. This is useful to know as the same DSM-IV ODD symptoms are used across different cultural groups for diagnosis (Canino et al., 2010). As the ODD symptom scores used here were binary scores, and as binary scores are categorical and correspond closely to the DSM-IV approach on how the ODD symptoms are scored for the diagnosis of ODD, the findings for equivalence for latent means scores can be taken as providing support for using the same-cut off scores for the diagnosis of ODD in the cultural groups examined in the study. At a general level this can be taken as adding some support for the use of the same cut-off scores for diagnosis of ODD across different cultural groups, as is usually followed in current clinical practice. However as this is highly speculative, this needs to be examined directly in future studies. Also, the cross-cultural equivalence suggested here for the ODD symptoms does not imply that cultural factors are irrelevant for understanding ODD. It may be that such factors are involved in understanding other relevant issues, such as societal responses to ODD in terms of general response and interventions.

While the findings in this and the Gomez et al. (2008) studies have been interpreted in terms of cross-cultural equivalence for the ODD symptoms, this interpretation is based on the assumption that the Malay, Chinese and Indian communities in Malaysia are culturally distinct. If this assumption is incorrect than this interpretation is also incorrect. There is a possibility for this. Malaysia is a progressive industrialized modern society, influenced by Western values and institutions (e.g., government, law,

medicine, science, politics, and education). Tu (2003) has proposed that as traditional non-Western communities modernize, their traditional values undergo a transformation that represents a mixture of both the traditional and modern Western values. Thus, it can be expected that both modern Western and ethnic-specific values exist in Malaysia, with Chinese, Malays and Indians sharing the common set of Western modern values. It could be that this common set of core modern values moderates the influence of the ethnic-specific values. This means that in reality the different cultural groups are not as distinct as suspected. This also means that it may be misleading to interpret the findings here in terms of cross-cultural equivalence.

In concluding, the findings and interpretations made in this study needs to be viewed with several limitations in mind. First, as already noted, despite being cultural distinct, the Malay, Chinese and Indian communities may be holding a common set of core modern values that have stronger influence on the behaviors than their own ethnic-specific values and therefore equivalence rather than no equivalence is the expectation. Second, although teachers in a Malaysian population would have come from different racial groups and therefore potentially confound findings, it is argued that this need not be a concern as the previous study by Gomez et al. (2008) showed that ratings of the ODD symptoms are not influenced by respondents' racial backgrounds. Third, it is uncertain if different results would emerge from interview-based data. Unlike rating scales, clinical interviews provide opportunities for clinicians to deal with respondents' uncertainties when answering questions. Fourth, since ODD rating scales are used to facilitate the diagnosis of ODD, it would be useful to replicate this study with children diagnosed or with ODD. Given the limitations noted here, some may wish to consider the results of this study as preliminary.

Source of funding

This study received no funding.

Conflict of interest

The author declares no conflict of interest.

Contribution

The author was responsible for planning of the study, preparing of research protocol and final writing of the article.

Acknowledgements

The author wishes to thank the Malaysian Government, and all schools and parents who participated in the study.

References

- American Psychiatric Association, 2000. *Diagnostic and Statistical Manual of Mental Disorders – IV Text Revised*. Author, Washington, DC.
- Barkley, R.A., Murphy, K.R., 1998. *Attention-deficit hyperactivity disorder: a clinical workbook*, 2nd ed. Guilford Press, New York.
- Butcher, J.N., 1996. Translation and adaptation of the MMPI-2 for international use. In: Butcher, J.N. (Ed.), *International Adaptations of the MMPI-2: Research and Clinical Applications*. University of Minnesota Press, Minneapolis, pp. 26–43.
- Canino, G., Polanczyk, G., Bauermeister, J.J., Rohde, I., Frick, A.P.J., 2010. Does the prevalence of CD and ODD vary across cultures? *Social Psychiatry and Psychiatric Epidemiology* 45, 695–704.
- Cohen, J., 1992. A power primer. *Psychological Bulletin* 112, 155–159.
- Department of Statistics Malaysia, 2003. *Population and Housing Census of Malaysia: Education and Social Characteristics of the Population*. Government of Malaysia, Kuala Lumpur.
- Gomez, R., Burns, G.L., Walsh, J.A., 2008. Parent ratings of the ODD symptoms: IRT analyses of cross-national and cross-racial invariance. *Journal of Psychopathology and Behavioral Assessment* 30, 10–19.
- Hofstede, G., 1980. *Culture's Consequences: International Differences in Work-related Values*. Sage, Beverly Hills, CA.
- Hu, L.T., Bentler, P.M., 1998. Fit indices in covariance structure modeling: sensitivity to under parameterized model misspecification. *Psychological Methods* 3, 424–453.
- Millsap, R.E., Yun-Tein, J., 2004. Assessing factorial invariance in ordered-categorical measures. *Multivariate Behavioral Research* 39, 479–511.
- Muthen, L.K., Muthen, B.O., 2010. *Mplus User's Guide*, 6th ed. Muthen & Muthen, Los Angeles, CA.
- Patterson, G.R., Reid, J.B., Dishion, T.J., 1992. *Antisocial Boys*. Castalia, Eugene, OR.
- Quah, S.R., 2004. Ethnicity and parenting styles among Singapore families. *Marriage & Family Review* 35, 63–83.
- Reise, S.P., Widaman, K.F., Paugh, P.H., 1993. Confirmatory factor analysis and item response theory: two approaches for exploring measurement invariance. *Psychological Bulletin* 114, 552–566.
- Stormshak, E.A., Bierman, K.L., McMahon, R.J., Lengua, L.J., 2000. Parenting practices and child disruptive behavior problems in early elementary school. *Journal of Clinical Child Psychology* 29, 17–29.
- Tu, W.M., 2003. Multiple modernities: a preliminary inquiry into the implications of East Asian modernity. In: Harrison, L.E., Huntington, S.P. (Eds.), *Culture Matters: How Value Shape Human Process*. Basic Books, New York, pp. 256–267.